

HNO 2011  
 DOI 10.1007/s00106-010-2231-6  
 © Springer-Verlag 2010

#### Redaktion

B. Wollenberg, Lübeck



K. Neumann<sup>1</sup> · I. Holler-Zittlau<sup>1</sup> · S. van Minnen<sup>2</sup> · U. Sick<sup>1</sup> · Y. Zaretsky<sup>1,3</sup> · H.A. Euler<sup>1</sup>  
<sup>1</sup> Schwerpunkt für Phoniatrie und Pädaudiologie, Klinik für Hals-Nasen-Ohren-Heilkunde, Goethe-Universität, Frankfurt am Main  
<sup>2</sup> Institut für Heil- und Sonderpädagogik, FB 03 Abteilung Sprachheilpädagogik, Universität Gießen  
<sup>3</sup> Institut für Medizinische Psychologie, Goethe-Universität, Frankfurt am Main

# Katzengoldstandards in der Sprachstandserfassung

## Sensitivität und Spezifität des Kindersprachscreenings (KiSS)

Sprachstandserfassungen im Kindergartenalter als flächendeckende Screeningprogramme rücken seit etlichen Jahren in vielen Industrieländern in den Blickpunkt des Interesses [36]. Auch in Deutschland wurde es bildungs- und gesellschaftspolitisches Ziel, durch kompetenzorientierte Sprachförderung oder – wenn nötig – Sprachtherapie eine ausreichende Deutschsprachfähigkeit der Schulanfänger zu sichern. Voraussetzung für individuell angepasste Förder- und Therapiekonzepte ist eine Erfassung des Sprachstands und seine Verlaufskontrolle [41, 48].

Die Prävalenz von Sprachentwicklungsstörungen (SES) liegt so hoch, dass allein sie ein Screening rechtfertigt, wobei dieses auch andere Sprachauffälligkeiten abdecken sollte. Eine SES liegt bei erheblichen zeitlichen und inhaltlichen Abweichungen nach unten von der normalen Sprachentwicklung im Kindesalter vor. Die Abweichungen können Sprachproduktion und/oder Sprachverständnis auf einer oder mehreren linguistischen Ebenen (phonetisch-phonologisch, lexikalisch-semantisch, morphologisch-syntaktisch, pragmatisch-kommunikativ) betreffen [42]. Prävalenzangaben zu SES schwanken auch für den gut beforschten angelsächsischen Sprachraum stark, werden aber am häufigsten zwischen 6 und

8% angegeben, wobei Jungen etwa doppelt so häufig betroffen sind wie Mädchen [13, 50]. Für Deutschland sind Prävalenzangaben uneinheitlicher und weniger belastbar wegen uneinheitlicher Bezugspopulationen [30, 54].

SES imponieren entweder isoliert als spezifische Sprachentwicklungsstörungen (SSES; „specific language impairment“, SLI) oder assoziiert mit weiteren Störungen bzw. Primärerkrankungen, z. B. Intelligenzminderung oder Hörstörung. Die hochgradig erblichen [9] SSES bilden die Teilmenge von SES, für die sprachrelevante Komorbiditäten nicht identifiziert werden können, und werden daher per Ausschluss diagnostiziert.

Nach der Internationalen Klassifikation der Erkrankungen (ICD-10, [16]) der Weltgesundheitsorganisation und ähnlich auch im häufig parallel verwendeten Diagnostischen und Statistischen Manual Psychischer Störungen (DSM-IV-TR, [2]) werden SSES im Wesentlichen unter „Umschriebene Entwicklungsstörungen des Sprechens und der Sprache“ (UESS, 80.-) gefasst. Allerdings gibt es einige Abweichungen zwischen beiden Entitäten. So beinhaltet F.80 auch phonetische Störungen und auditive Verarbeitungs- und Wahrnehmungsstörungen (AVWS), die gar nicht bzw. nur teilweise zu den SSES gezählt werden dürfen [41]. Da für AV-

WS im deutschen Sprachraum noch keine ausreichend validierten und standardisierten Testbatterien empfohlen werden können, werden sie aus der weiteren Betrachtung herausgelassen. Aus Praktikabilitätsgründen werden UESS und SSES im Folgenden synonym verwendet.

Die Prävalenz von UESS im Kindesalter wird für den amerikanischen Sprachraum nach den ICD-Kriterien mit 5–8% angegeben [2, 52] und auch für den deutschen Sprachraum so erwartet, auch wenn hier uneinheitliche Angaben und Klassifikationskriterien existieren [30, 32, 54]. Eine UESS beeinträchtigt die schulische und Gesamtentwicklung eines Kindes nachhaltig [3, 4, 6, 7, 8, 11, 49], bildet sich ohne Intervention meist nicht vollständig zurück und bedarf einer fachspezifischen Diagnostik und Behandlung. Die Symptome können bis in das Erwachsenenalter persistieren [21] und den beruflichen Werdegang behindern [14, 15, 18]. Etwa 40–75% der Kinder mit UESS haben später Probleme im Schriftspracherwerb.

Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) wurde 2006 vom Gemeinsamen Bundesausschuss beauftragt, den Nutzen einer Früherkennungsuntersuchung auf UESS bei Kindern bis zu 6 Jahren zu bewerten. Im Abschlussbericht des IQWiG wurde die Zielstellung früher Sprach-

screenings allerdings ausgeweitet auf die Identifikation von Kindern mit einem besonders hohen Risiko „... für Schwierigkeiten in der Schulausbildung sowie für Beeinträchtigungen im zwischenmenschlichen und emotionalen Bereich ...“ ([28]: S. V), also deutlich über UESS hinausgehend. Es wird resümiert, dass derzeit der Nutzen einer solchen Früherkennungsuntersuchung nicht belegt sei, was aber nicht bedeute, dass es ihn nicht gäbe. Für eine Prüfung der Erfüllung notwendiger Voraussetzungen für ein Screeningprogramm sei jedoch zu belegen, dass Kinder mit Therapiebedarf ausreichend zuverlässig diagnostiziert werden könnten, dass nicht zu viele Kinder fälschlicherweise als behandlungsbedürftig eingestuft würden und dass die Behandlung von Sprachstörungen im Vorschulalter der Behandlung im Schulalter überlegen sei. Es gäbe derzeit keinen Beleg dafür, dass Sprachtherapien bei jüngeren Kindern einen höheren Nutzen hätten als bei älteren und dass die Wirkung von Sprachtherapien über kurzfristige positive Effekte hinausginge. Zudem seien für deutschsprachige Diagnostikinstrumente keine Gütekriterien bezogen auf die Entdeckung von Kindern mit UESS in der Gesamtbevölkerung vorhanden. Somit fehlten in Deutschland derzeit wesentliche methodische Grundlagen für ein Screening von UESS. Sollten diese jedoch geschaffen werden, könne man den potenziellen Nutzen und Schaden eines Screenings untersuchen [28].

Das IQWiG bewertet allerdings medizinische Prozeduren ausschließlich unter dem Aspekt der Evidenzbasiertheit. Ein hohes Evidenzniveau für Verfahren zur Identifikation, Diagnostik und Therapie von SES zu erbringen, ist jedoch wegen der großen Variabilität der Sprachentwicklung und ihrer Störungsbilder schwieriger als für andere Krankheitsbilder, die beispielsweise durch Laboruntersuchungen oder apparative Tests zu diagnostizieren und medikamentös zu behandeln sind. Dennoch ist anzustreben, streng algorithmisierte Verfahren mit nachgewiesenen psychometrischen Gütekriterien zu entwickeln und einzusetzen. Dies ist auch Ziel der hier vorgestellten Studie.

Sprachscreenings sollten nicht nur zur frühen Identifikation von Kindern mit

UESS dienen, sondern auch zur Auffindung von Spracherwerbsdefiziten überhaupt, um die betroffenen Kinder entweder einer individuellen Sprachförderung oder – bei Verdacht auf SES – einer Diagnostik und ggf. Sprachtherapie zuzuführen. Ersteres betrifft v. a. Kinder aus bildungsfernen Milieu oder mit Migrationshintergrund. Unbedingt und unschwer ist ein Sprachscreening auch auszuweiten auf Redefluss- (Stottern, Poltern) und Stimmstörungen, die im Kindergartenalter Prävalenzen von 3–5% haben und bei früher Therapie eine erheblich bessere Prognose aufweisen als bei späterer [33, 47, 56].

Einige Voraussetzungen für ein Screening, wie substanzielle Prävalenzrate oder ungünstige Langzeitprognose bei Nichtbehandlung, sind gegeben [38], andere, z. B. eine erfolgversprechende Sprachförderung oder -therapie, sind, wie auch im IQWiG-Bericht erwähnt, nicht für alle Arten kindlicher Sprachdefizite zweifelsfrei nachgewiesen [28, 35, 55]. Allerdings bergen derzeit vorhandene Sprachscreenings einige Probleme. In der Regel erfassen sie nur bestimmte sprachliche Fähigkeiten, z. B. durch Nachsprehaufgaben [12, 25]. Psychometrische Gütekriterien als obligatorische Voraussetzungen für die Bestimmung des Anwendungsnutzens sind nicht immer hinreichend belegt. Das gilt insbesondere für nichtstandardisierte Beobachtungen, denen oft eine Siebtstgüte oder gar diagnostische Güte unkritisch zugetraut wird.

Der Mangel an objektiven Messverfahren und die große Variabilität in der Lautsprachentwicklung erschweren die Feststellung von Sprachauffälligkeiten. Sensitivitäten und Spezifitäten eines Screenings auf UESS von je 80% werden daher als akzeptabel erachtet [44]. In einer US-amerikanischen Übersicht über Sprachscreenings für Kinder im Alter von 3–5 Jahren werden 12 Studien aufgeführt, davon 6 mit ausreichender und eine mit guter methodischer Qualität [53]. Bei diesen 7 akzeptablen Studien lag die Sensitivität zwischen 60 und 95% (Median: 66%) und die Spezifität zwischen 48 und 100% (Median: 90%). Auch wenn also im angloamerikanischen Sprachraum einige Sprachscreenings die Gütekriterien erfüllen, wird dennoch die Evidenz für den Nutzen flächendeckender Screenings als nicht

ausreichend erachtet und ihre Sensitivität kritisch beurteilt, u. a. weil Kinder ohne sprachliche Defizite leichter zu identifizieren sind als Kinder mit Defiziten [34] und der langfristige Nutzen resultierender Interventionsmaßnahmen nicht hinreichend belegt ist [40].

Auch die hessische Landesregierung betreibt seit 2006 die Einführung einer flächendeckenden Sprachstandserfassung in den Kindertageseinrichtungen. Dafür wurde ein Instrument für vier- bis viereinhalbjährige Kinder entwickelt, das Kindersprachscreening (KiSS). Seine Konstruktion sowie Konstrukt- und psychometrische konkurrente Validität sind andernorts beschrieben [19]. Es wird von Erzieher(inne)n<sup>1</sup> in den Kindertagesstätten nach Schulung durch supervidierende Sprachexpertinnen (sprachtherapeutischer Beruf) durchgeführt, sowohl bei muttersprachlich deutschen als auch bei anderssprachigen Kindern. Es soll nicht nur sprachauffällige Kinder identifizieren, sondern auch unterscheiden, welche Kinder eine pädagogische Sprachförderung im Rahmen der Kindertageseinrichtung benötigen („sprachpädagogisch auffällig“) und welche Kinder einer weiteren klinischen Abklärung und ggf. Therapie bedürfen („klinisch auffällig“ bei vermuteten Sprachentwicklungs-, Redefluss- und Stimmstörungen; [19]).

Hier sollen

- die Ergebnisse der an 257 Kindern durchgeführten Validierungsstudie des KiSS beschrieben,
- eine Abschätzung der nach flächendeckender Implementierung des Screenings zu erwartenden sprachlichen Auffälligkeiten vorgenommen,
- seine klassifikatorischen Gütekriterien dargestellt und
- die Eignung des KiSS speziell zur Auffindung von schulerfolgsrelevanten SES bzw. UESS entsprechend den Kriterien des IQWiG-Auftrags überprüft werden.

Wegen möglicher Abgrenzungsschwierigkeiten von Störungsbildern ist eine terminologische Klärung vonnöten. Als

<sup>1</sup> Zur besseren Lesbarkeit wird nur eine Geschlechtsform verwendet; gemeint sind immer beide.

HNO 2011 · [jvn]:[afp]–[alp] DOI 10.1007/s00106-010-2231-6  
© Springer-Verlag 2010

K. Neumann · I. Holler-Zittlau · S. van Minnen · U. Sick · Y. Zaretsky · H.A. Euler

### Katzengoldstandards in der Sprachstandserfassung. Sensitivität und Spezifität des Kindersprachscreenings (KiSS)

#### Zusammenfassung

**Hintergrund.** Für eine flächendeckende Sprachstandserfassung in Hessen wurde das Kindersprachscreening (KiSS) entwickelt zur Identifikation von Kindern im Kindergartenalter mit sprachpädagogischem Förderbedarf und mit klinischem Abklärungsbedarf.

**Teilnehmer und Methoden.** Zur Ermittlung der klassifikatorischen Verfahrensvalidität durchliefen 257 Kindergartenkinder (Alter 4;0–4;5 Jahre) Untersuchungen mit dem KiSS sowie den Sprachtests Reynell-Sprachentwicklungsskalen, Patholinguistische Diagnostik, Psycholinguistische Analyse kindlicher Sprechstörungen (PLAKSS) und Aktiver Wortschatztest für 3- bis 5-jährige Kinder – Revision (AWST-R). Als Referenzkriterium für den Auffälligkeitsbefund des KiSS diente das Majoritäts- bzw. Konsensurteil dreier ausgewie-

sener Sprachexpertinnen, denen die Ergebnisse der Sprachentwicklungstests vorlagen. Die Grundraten/Auffälligkeitsraten der Expertinnen waren entweder selbstbestimmt oder prävalenzbasiert vorgegeben.

**Ergebnisse.** Für Letztere wurden wegen besserer Beurteilerübereinstimmung höhere Validitätswerte als für selbstbestimmte Grundraten ermittelt. Für die Kreuztabellierung der im KiSS insgesamt Auffälligen (sprachpädagogisch oder klinisch) mit dem Expertenurteil zu Sprachentwicklungsstörungen (SES) einschließlich Redefluss- und Stimmstörungen (bei vorgegebener Grundrate) waren die Sensitivität 88% und die Spezifität 78%, für Sprachentwicklungsstörungen allein 84% und 75%. Die Spezifitäten für die klinischen Auffälligkeiten im KiSS (SES allein oder kom-

biniiert mit Redefluss- und Stimmstörungen) bezogen auf das testbasierte Konsens-Expertenurteil lagen bei 93%, die Sensitivitäten waren unbefriedigend, da die Unterscheidung zwischen sprachpädagogischer und klinischer Auffälligkeit noch verbesserungsbedürftig war.

**Schlussfolgerung.** Da die Konkordanz der Sprachexpertinnen nur mäßig war, scheint die Entwicklung eines deutschsprachigen umfassenden Referenztests für SES mit evidenzbasierten algorithmischen Entscheidungsregeln erforderlich.

#### Schlüsselwörter

Sprachscreening · Sprachstörungen · Sprachentwicklungsstörungen · Massenscreening · Sensitivität und Spezifität

### Fool's gold standards in language screening. Sensitivity and specificity of the Hessian child language screening test (Kindersprachscreening, KiSS)

#### Abstract

**Background.** The German *Kindersprachscreening* (KiSS) is a universal speech and language screening test for large-scale identification of Hessian kindergarten children requiring special educational language training or clinical speech/language therapy.

**Participants and Methods.** To calculate the procedural screening validity, 257 children (aged 4.0 to 4.5 years) were tested using KiSS and four language tests (Reynell Development Language Scales III, Patholinguistische Diagnostik, PLAKSS, AWST-R). The majority or consensus judgements of three speech-language professionals, based on the language test results, served as a reference criterion. The base (fail) rates of the professionals were either self-determined or preset based on known prevalence rates.

**Results.** Screening validity was higher for preset than for self-determined base rates due to higher inter-judge agreement. The confusion matrices of the overall index classification of the KiSS (speech-language abnormalities with educational or clinical needs) with the fixed base rate expert judgement about language impairment, including fluency or voice disorders, yielded a sensitivity of 88% and a specificity of 78%, for just language impairment 84% and 75%, respectively. Specificities for disorders requiring clinical diagnostics in the KiSS (language impairment alone or combined with fluency/voice disorders) related to the test-based consensus expert judgment was about 93%. Sensitivities were unsatisfactory because the differentia-

tion between educational and clinical needs requires improvement.

**Conclusion.** Since the judgement concordances between the speech-language professionals was only moderate, the development of a comprehensive German reference test for speech and language disorders with evidence-based algorithmic decision rules rather than subjective clinical judgement is advocated.

#### Keywords

Language tests · Language disorders · Language development disorders · Mass screening · Sensitivity and specificity

sprachpädagogisch auffällig gelten Kinder mit einem Förderbedarf in der Kindertagesstätte, als *klinisch auffällig* solche mit medizinischem Abklärungs- und ggf. Therapiebedarf. Die klinische Auffälligkeit wird als *exklusiv* bezeichnet, wenn nur der Verdacht auf SES bewertet wird und als *inklusiv*, wenn außerdem Redefluss- und Stimmstörungen einbezogen sind. Da *Gesamtaufälligkeit* sprachpädagogische und/oder klinische Auffälligkeit einschließt, kann die Gesamtaufälligkeit ebenfalls exklusiv (ohne Redefluss- und Stimmstörungen) oder inklusiv sein. Die exklusive klinische Auffälligkeit entspricht in relativ guter Annäherung den UESS nach ICD-10 F80.-. Allerdings beinhaltet sie zusätzlich die mit Komorbiditäten assoziierten SES [41], allerdings nicht Kinder mit isolierten auditiven Verarbeitungs- und Wahrnehmungsstörungen.

## Methode

### Studienteilnehmer

Mit 257 unselektierten Kindergartenkindern im Alter von 4;0 bis 4;5 Jahren (137 männlich, 120 weiblich) aus 43 Kindertagesstätten in Frankfurt, Offenbach, Gießen, Landkreis Fulda und Landkreis Korbach wurden das revidierte Marburger Sprach-Screening für 4- bis 6-jährige Kinder (MSS, [27]), das die Grundlage des KiSS bildete und die später ausgewählten KiSS-Items enthält, sowie eine Reihe weiterer Sprachtests und -screenings durchgeführt. Bei 77 Kindern (30%) bestand ein Migrationshintergrund.

### Kindersprachscreening

Das KiSS umfasst folgende 3 Bestandteile:

- Kinderbogen mit Bildvorlage (Screeningbogen),
- Elternbogen (Fragebogen für Eltern) mit 25 Items und
- Kitabogen (Fragebogen für Erzieher/Erzieherinnen) mit 26 Items.

Es ist einsehbar unter der Webadresse „<http://www.hmafg.hessen.de/> >Gesundheit >Kindersprachscreening KiSS. Das Screening besteht aus 11 Untertests:

- Pragmatik (Spontanäußerung zum Bild);
- Sprachverständnis;
- Sprachproduktion;
- Aussprache;
- Wortschatz, sowie Wort- und Satzgrammatik anhand von
- Pluralbildung,
- Partizipbildung,
- Subjekt-Verb-Kongruenz bei der 2. Person Singular,
- Präpositionen im Akkusativkontext und
- im Dativkontext, und
- Nebensatzbildung (Details s. [19]).

Die Eltern- und Kitabögen wurden zur Erhebung anamnestisch bedeutsamer Befunde verwendet wie Familiensprache, Komorbiditäten, familiäre Sprachstörungen, Kommunikationsverhalten, Redefluss- und Stimmstörungen, bisherige oder gegenwärtige Therapien. Bei Diskrepanz zwischen Testergebnis und Fragebogenbefund oder bei wichtigen Hinweisen aus den Fragebögen, z. B. auf eine Hörstörung, wurde eine Nachbeurteilung des Kindes durch eine Sprachtherapeutin aufgrund eines itemspezifischen Algorithmus veranlasst.

### Algorithmische Definition der Trennstellen

**Sprachpädagogische Auffälligkeit.** Ein monolingual deutschsprachiges Kind wird im KiSS als „sprachpädagogisch auffällig“ eingestuft, wenn es mindestens eines der folgenden Kriterien nicht erreichte:

- 48 Punkte im KiSS-Gesamtwert<sup>2</sup>,
- spontane Äußerung zum Bild oder Äußerung auf Nachfrage,
- 2 Punkte im Untertest Sprachverständnis,
- 9 Punkte im Untertest Wortschatz,
- 7 Punkte im Untertest Lautbildung,
- 4 Punkte im Untertest Grammatik.

Ein Kind mit nichtdeutscher Muttersprache wird als sprachpädagogisch auffällig

bewertet, wenn mindestens eines der folgenden Kriterien nicht erreicht wurde:

- 35 Punkte im KiSS-Gesamtwert,
- spontane Äußerung zum Bild oder Äußerung auf Nachfrage,
- 2 Punkte im Untertest Sprachverständnis,
- 6 Punkte im Untertest Wortschatz,
- 7 Punkte im Untertest Lautbildung,
- 4 Punkte im Untertest Grammatik.

**Klinische Auffälligkeit.** Ein Kind wird unabhängig von seinem Sprachhintergrund als „inklusiv klinisch auffällig“ bewertet, wenn es mindestens eines der folgenden Kriterien nicht erreichte:

- 6 Punkte im Untertest Lautbildung,
- Aussprache des Lauts /k/ bei mindestens einem von 4 Testwörtern korrekt,
- Aussprache des Lauts /r/ bei mindestens einem von 4 Testwörtern korrekt,
- keine Redeflussstörung (Stottern, Poltern) und
- keine häufige oder dauerhafte Stimmstörung nach Auskunft der Erzieherin.

Muttersprachlich deutsche Kinder mussten zudem

- mindestens 4 Punkte im Untertest Grammatik aufweisen. Die Einstufung als „exklusiv klinisch auffällig“ lässt die Kriterien der Redeflussstörung und der Stimmstörung unbeachtet.

Da in der untersuchten Altersklasse eine Grammatikstörung als eines der vorherrschenden Störungsbilder einer SES gilt, bestand bei der Erstellung des Algorithmus für Kinder mit Migrationshintergrund und nichtdeutscher Herkunftssprache die Schwierigkeit, dass die grammatischen Leistungen einerseits wegen mangelnder Deutschkenntnisse eingeschränkt sein können, andererseits aber auch oder zusätzlich wegen einer SES. Daher erhalten Migrantenkinder mit weniger als 4 Punkten im Untertest Grammatik, die bereits länger als ein Jahr eine Kindertagesstätte besuchen, eine alsbaldige Sprachförderung und werden nach 6 Monaten durch eine Sprachexpertin mit dem KiSS nachgetestet. Wird ein Entwicklungsfortschritt von mindestens 8 Punkten KiSS-Gesamtscore überschrit-

<sup>2</sup> In der statistischen Analyse wurde mit Mittelwerten statt mit Rohwerten gerechnet, um den listenweisen Ausschluss von Fällen durch fehlende individuelle Einzelangaben zu verhindern.

ten, so wird von der Sprachexpertin entschieden, ob eine weitere Förderung nötig ist oder nicht. Bleibt das Kind unter diesem Zuwachs oder erreicht es weniger als 5 Punkte im Untertest Grammatik, wird eine medizinische Abklärung empfohlen. Eine nicht regelhafte Zischlautbildung, die im Untersuchungsalter noch nicht abgeschlossen sein muss, führt zur Empfehlung einer Kontrolluntersuchung im Alter von 5 Jahren [19].

## Referenzverfahren und zusätzliche Screenings

Die Zusammenstellung von geeigneten Verfahren für eine Referenztestbatterie verlangt Qualitätskompromisse. Umfassende deutsche Sprachtests für die hier untersuchte Altersklasse verfügen entweder nicht über aktuelle Normwerte oder sind zu stark intelligenzabhängig. Andere Verfahren untersuchen nur ausgewählte linguistische Ebenen und unterscheiden sich in ihrer Testgüte, sofern Gütekriterien überhaupt beschrieben und bezifferbar sind [41]. Damit ist im deutschsprachigen Raum ein Goldstandard als externes (konkurrentes) Validierungskriterium und Bewertungsmaßstab der klassifikatorischen Validität eines Sprachstandscreenings im Vorschulalter nicht als einzelner, umfassender Test verfügbar, sondern muss aus mehreren Kriterien zusammengestellt werden. Daher wurde für die Bestimmung der klassifikatorischen Validität des KiSS ein auf den Kennwerten von 4 Referenztests basiertes Expertenurteil herangezogen.

Als Referenztests dienen

- der rezeptive Teil der Reynell Development Language Scales III ([46], aktuelle 3. deutsche Version) für die Untersuchung des Sprachverständnisses,
- die Psycholinguistische Analyse kindlicher Sprechstörungen (PLAKSS, [23]) für die Bewertung der Aussprache,
- die Untertests zur Produktion von Satzstrukturen, Nominalphrasen, Dativ-, Akkusativ- und Pluralmarkierungen aus der Patholinguistischen Diagnostik bei Sprachentwicklungsstörungen (PD, [29]) zur Einschätzung der Grammatikleistungen und

- der Aktive Wortschatztest für 3- bis 5-jährige Kinder – Revision (AWST-R, [31]) zur Bewertung der semantisch-lexikalischen Fähigkeiten.

Vergleichend wurden 2 psychometrisch konstruierte Screenings mit belegbaren Gütekriterien einbezogen, das Sprachscreening für Kinder im Vorschulalter (SSV, [25]) und das Heidelberger Auditive Screening in der Einschulungsuntersuchung (HASE, [12]). Beide Tests überprüfen das phonologische Arbeitsgedächtnis über das Nachsprechen von Sätzen und von Nonsens- bzw. Zahlwörtern. Sie untersuchen damit die phonologische Speicherkomponente des auditiven Arbeitsgedächtnisses [5], die es dem Kind ermöglicht, größere, noch unanalyisierte Einheiten im phonologischen Arbeitsgedächtnis zur Verfügung zu haben, was in kausaler Beziehung zu Laut- und Wortschatzerwerb, Grammatikentwicklung und Schriftspracherwerb steht [25, 26].

## Durchführung

Jedes Kind wurde in der Kindertageseinrichtung mit dem revidierten MSS, dem SSV und dem HASE entweder durch die von Sprachexpertinnen in der Durchführung geschulten Erzieherinnen oder durch Sprachexpertinnen getestet. Letztere führten auch die Referenztests durch. Jedes Kind nahm an 4–5 Sitzungen von je etwa einer Stunde teil [19].

## Findung des Referenzkriteriums

Die Testwerte des KiSS korrelierten hoch [19] mit den aggregierten z-transformierten Testwerten aus den genannten 4 gängigen Tests zur Erfassung der Sprachkompetenz (Reynell, PLAKSS, PD, AWST-R). Damit weist das Verfahren eine hohe konkurrente psychometrische Kriteriumsvalidität auf. Das Screening verlangt jedoch als Resultat eine kategoriale Entscheidung darüber, ob eine Auffälligkeit vorliegt, die weitere Maßnahmen (z. B. Sprachförderung, Abklärungsdiagnostik und ggf. Therapie) nahelegt, oder ob das Kind als unauffällig eingestuft werden darf.

In der Signalentdeckungstheorie als formaler Grundlage der Gütekriterienbewertung von Screenings [24] wird vo-

rausgesetzt, dass das Vorhandensein oder Fehlen des Signals fehlerfrei bekannt ist. Diese Voraussetzung ist jedoch bei einem Sprachscreening nicht gegeben. Es existiert derzeit kein unangefochtenes, deutschsprachiges, zeitgemäßes, alle linguistischen Ebenen valide umfassendes Diagnostikverfahren zur Feststellung, ob eine Sprachentwicklungsstörung mit entsprechenden Konsequenzen für den weiteren oralen und literalen Spracherwerb wirklich vorliegt. Allein die Tatsache, dass in der deutschsprachigen Literatur die Angaben über die Prävalenzen für SES von niedrig einstelligen bis weit in zweistellige Prozentangaben variieren [28], spricht für mangelnde Objektivität (i. S. von Intersubjektivität) ihrer Diagnose, und die Objektivität eines Tests definiert die Obergrenze seiner erreichbaren Validität.

Die Referenztests selbst sind als Kriterium für eine dichotome Einstufung nicht durchweg geeignet. Nur 2 der 4 Referenztests (Reynell, AWST-R) liefern Altersnormen, die eine nachvollziehbare algorithmische Klassifikation (auffällig – unauffällig) ermöglichen. Die Auffälligkeitskriterien für PD und PLAKSS scheinen nicht durchgehend statistisch normenbasiert zu sein. Für die PLAKSS beispielsweise ist Auffälligkeit angezeigt, wenn ein Drittel der Items nicht richtig beantwortet wird. Gibt man sich dennoch mit den für die Referenztests gegebenen Kriterien zufrieden und definiert ein Gesamtaufälligkeitskriterium aus der Addition der relationalen Operatoren (d. h. auffällig, wenn in mindestens einem der Referenztests auffällig), dann würde man die Auffälligkeitshäufigkeiten addieren. Dieses Verfahren würde dazu führen, dass in unserer Stichprobe 66% der Kinder exklusiv klinisch auffällig, d. h. SES-verdächtig, wären, was einen solchen Klassifikationsalgorithmus wegen Überdiagnose disqualifiziert. Deswegen wurde ein Annäherungsverfahren an eine möglichst verlässliche Referenzdiagnose gewählt, das nun in den einzelnen Schritten beschrieben wird.

## Anteil sprachauffälliger Kinder bei Expertenurteil mit freier Grundrate

Drei in der Diagnostik und Behandlung von kindlichen Sprachentwicklungsstö-

rungen erfahrene und akademisch ausgewiesene Sprachexpertinnen wurden gebeten, den Sprachstand der Kinder auf Grundlage der Ergebnisse der Testbatterie zu beurteilen. Sie sollten nach ihrem subjektiven fachlichen Urteil bei der Interpretation der Testergebnisse unabhängig voneinander angeben, welche der 257 Kinder sprachpädagogisch, exklusiv klinisch oder inklusiv klinisch auffällig seien. Der Prozentanteil der von den Expertinnen als auffällig eingestuften Kinder (diagnosepositiv) wird hier als Grundrate, der Anteil der im KiSS auffälligen Kinder als Selektionsrate (testpositiv) bezeichnet. Den Expertinnen war die Wahl der Grundrate freigestellt.

Die 3 Expertinnen stuften 33%, 33% und 26% der Kinder als sprachpädagogisch auffällig ein, 25%, 22% und 28% als exklusiv klinisch auffällig, 25%, 26% und 35% als inklusiv klinisch auffällig und 38%, 39% und 50% als inklusiv gesamtauffällig. Angesichts der Prävalenzen von SES, Redefluss- und Stimmstörungen [37, 47, 56] scheinen diese Einschätzungen zu hoch und zwischen den Expertinnen zu stark zu variieren, um ohne Weiteres als externes Referenzkriterium dienen zu können. Deswegen sollten Vorgaben in Form von Empfehlungen zur Bildung von Rangabstufungen das Verfahren zur Beurteilerübereinstimmung ergänzen.

### Expertenübereinstimmung bei ordinaler Beurteilung der klinischen Auffälligkeit mit freier Grundrate

Die Beurteilerübereinstimmung wird außer durch inhaltliche Übereinstimmung auch durch subjektive Positionierungen der Cut-offs und damit durch individuell unterschiedliche Grundraten beeinflusst. Um die inhaltliche Konkordanz in Bezug auf klinische Auffälligkeit unter Minimierung des Einflusses von individuell variierenden Grundraten zu bestimmen, wurden die Expertinnen angewiesen, alle 257 Kinder nach dem Grad der inklusiven klinischen Auffälligkeit unabhängig voneinander ordinal einzustufen, von „sicher klinisch unauffällig“ bis „sicher klinisch auffällig“. Die inhaltlichen Kriterien für die Einstufungen waren wiederum freigestellt und sollten ohne gegenseitige Absprachen erfolgen. Die

Anzahl der zu verwendenden Rangabstufungen war freigestellt, sollte aber mindestens 3 betragen.

Diese Abstufungen sollten so gewählt werden, dass die 34 am meisten auffälligen Kinder entnehmbar waren, da auch das KiSS 34 inklusiv klinisch auffällige Kinder in der untersuchten Population ermittelt hatte. Zwischen dem 34. Kind und dem 35. Kind sollte sich also eine Rangstufe befinden. Dadurch wurde gewährleistet, dass alle 3 Expertinnen die gleiche Anzahl von inklusiv klinisch sprachauffälligen Kindern identifizierten (14%). Bei der Festlegung des Cut-offs für inklusiv klinische Auffälligkeit im KiSS hatten sich die Testautoren an den Prävalenzraten für SES, Redefluss- und Stimmstörungen orientiert und eine gewisse Überlapung der Störungsbilder sowie einen angemessenen Anteil falsch-positiver Screeningergebnisse einkalkuliert, sodass eine Auffälligkeitsrate zwischen 13 und 15% erwartet werden konnte.

Die Expertinnen benutzten jeweils 8 Rangstufen für inklusiv klinische Auffälligkeit. Der Kendall-Konkordanzkoeffizient für die Beurteilerübereinstimmung [10] war  $W=0,63$ , was eine hinreichende inhaltliche Übereinstimmung der Expertinnen in Bezug auf inklusiv klinische Auffälligkeit belegt, um bei einheitlichen Grundraten und optimiertem Cut-off-Kriterium als externe Referenz für die Bestimmung von Sensitivität und Spezifität des KiSS dienen zu können.

### Expertenübereinstimmung bei dichotomer Beurteilung der klinischen Auffälligkeit

**Festgelegte Grundrate.** Da eine inhaltliche Übereinstimmung in Bezug auf inklusive klinische Auffälligkeit gegeben war, die Cut-offs aber beurteilerspezifisch waren, wurde die Übereinstimmung zwischen den Beurteilerinnen bei vorgegebener Grundrate (34 von 257 Kindern, 13,6%) bestimmt.

Bei der Quantifizierung der Urteilskonkordanz für mehrere Beurteiler bei binären Daten findet das Nominalskalen- $\kappa$  von Fleiss [10, 22] Anwendung. Dieser  $\kappa$ -Koeffizient bestimmt den Anteil an konkordanten Urteilsparen aus allen möglichen Urteilsparungen, die über Zufalls-

übereinstimmungen hinausgehen. Hier gibt es für jedes Kind 3 Beurteilungen nach „klinisch auffällig“ oder „klinisch unauffällig“ und damit 3 mögliche Urteilsparungen. Von den bei 257 Kindern 771 insgesamt möglichen Urteilsparungen stimmten 87,6% überein. Da 34 von 257 Kindern binär zu beurteilen waren, ergibt sich eine Zufallsübereinstimmung von 64,8%. Der aus beiden Prozentzahlen errechnete  $\kappa$ -Koeffizient ist  $\kappa_m=0,65$ . Er stellt den relativen Anteil der überzufälligen Konkordanz an den maximal möglichen Konkordanz dar. Für den  $\kappa$ -Koeffizienten von Fleiss lässt sich bei bestimmten Stichprobenmindestumfängen, die hier gegeben sind, die Zufallswahrscheinlichkeit über die Prüfstatistik  $u$ , die standardnormalverteilt ist, berechnen ([10], S. 457). Es ergab sich ein  $u=15,8$ , was bei einseitiger Fragestellung einer Signifikanz von  $p<0,001$  entspricht. Die Konkordanz bei festgelegter Grundrate für inklusiv klinische Sprachauffälligkeit war befriedigend.

Die Expertinnen sollten zudem die 27 Kinder identifizieren, die mit größter Wahrscheinlichkeit exklusiv klinisch auffällig waren (SES-Verdacht). Dieses Vorgehen wurde gewählt, da auch das KiSS 27 exklusiv klinisch auffällige Kinder ermittelt hatte (10,5%). Die Zufallsübereinstimmung ist hier 81,2%, die tatsächliche Übereinstimmung war 89,6%, was einen  $\kappa$ -Koeffizienten von  $\kappa_m=0,45$  ergibt ( $u=4,49$ ;  $p<0,001$ ; einseitig). Auch die exklusive klinische Auffälligkeit wurde von den Expertinnen befriedigend konkordant beurteilt.

### Majoritäts- und Konsensurteil

Da die Beurteilerübereinstimmungen für exklusive klinische Auffälligkeit ( $\kappa_m=0,45$ ), inklusive klinische Auffälligkeit ( $\kappa_m=0,65$ ), sprachpädagogische Auffälligkeit ( $\kappa_m=0,57$ ), exklusive Gesamtauffälligkeit ( $\kappa_m=0,87$ ) und inklusive Gesamtauffälligkeit ( $\kappa_m=0,65$ ) insgesamt befriedigend waren, wurde als endgültiges Referenzkriterium für die verschiedenen Arten von Testauffälligkeit das Majoritätsurteil bestimmt: Es galt diejenige Einstufung, die bei mindestens 2 der 3 Expertinnen übereinstimmte. Bei der Identifikation der 34 Kinder, die am deutlichsten inklusiv klinisch auffällig waren, wur-

de zusätzlich das Konsensurteil bestimmt: Die Expertinnen einigten sich nach eingehender inhaltlicher Diskussion auf diese Fälle. Das Konsensurteil mag möglicherweise näher am „wahren“ Sachverhalt als das Majoritätsurteil sein, aber es ist aufwendiger. Für das Majoritätsurteil ist ein 3. Urteil nur dann einzuholen, wenn die ersten beiden Beurteilungen nicht übereinstimmen.

### Ermittlung der Screeninggütekriterien des KiSS

Aus dem Ergebnis des KiSS und dem Majoritäts- bzw. Konsensurteil der Sprachexpertinnen wurden für die verschiedenen Kategorien von Sprachauffälligkeit Kreuztabellen und wichtige kategoriale Gütekriterien eines Screenings ermittelt (Tab. 8, 11), z. B. Sensitivität, Spezifität und RATZ-Index. Der RATZ-Index, ein im deutschen Sprachraum häufig benutztes Gütemaß von Sprachstandscreenings, gibt den relativen Anstieg der Trefferquote über die Zufallstrefferquote an, unter Berücksichtigung der maximal möglichen Trefferquote, die nur bei gleicher Grund- und Selektionsrate 100% sein kann [20, 38]. Im Unterschied zu dem vergleichbaren  $\kappa$ -Koeffizienten  $\kappa_m$  nach Fleiss hat der RATZ-Index den Nachteil, dass er nicht auf Signifikanz überprüfbar ist. Stattdessen hat es sich eingebürgert, RATZ-Indizes unter 33% als unzureichende prognostische Validität zu werten, Indizes über 33% als gut, wenn auch unspezifisch, und Indizes über 66% als sehr gut [39]. Die statistische Analyse wurde mit BiAS, Version 8.6, [1] und SPSS 17.0 ausgeführt.

### Ergebnisse

Die Screeningergebnisse finden sich in Tab. 1. Demnach wurden mit dem KiSS 30,8% inklusiv gesamtauffällige Kinder identifiziert. Inklusiv klinisch auffällig waren 14,2%, exklusiv 11,1% (SES-Verdacht), und die Untersucher erkannten Stimmstörungen bei 0,4% und Redeflussstörungen bei 4,1% der Kinder. Sprachpädagogisch auffällig waren 25,9% der Kinder. Erwartungsgemäß waren hier Migrantenkinder erheblich stärker vertreten (45,2%) als monolingual deutsche Kinder (17,6%). Die 34 Kinder mit inklusiver

**Tab. 1** Häufigkeiten von auffälligen Kindern in den verschiedenen Klassifikationskategorien des KiSS

Auffälligkeit im KiSS	Muttersprache deutsch		Migrantenkinder		Alle Kinder	
	n	Auffällig n (Anteil in %)	n	Auffällig n (Anteil in %)	n <sup>a</sup>	Auffällig n (Anteil in %)
Inklusiv gesamtauffällig	168	38 (22,6)	72	36 (50,0)	240	74 (30,8)
Sprachpädagogisch auffällig	170	30 (17,6)	73	33 (45,2)	243	63 (25,9)
Inklusiv klinisch auffällig	169	24 (14,2)	71	10 (14,1)	240	34 (14,2)
Inklusiv klinisch und sprachpädagogisch auffällig	168	16 (9,5)	71	8 (11,3)	239	24 (10,0)
Exklusiv gesamtauffällig	169	34 (20,1)	74	34 (45,9)	243	68 (28,0)
Inklusiv klinisch auffällig	170	20 (11,8)	73	7 (9,6)	243	27 (11,1)
Exklusiv klinisch und sprachpädagogisch auffällig	169	16 (9,5)	73	7 (9,6)	242	23 (9,5)

<sup>a</sup> n kann wegen fehlender Daten variieren.

**Tab. 2** Häufigkeiten von Auffälligkeiten im KiSS

Testleistungen im KiSS	Muttersprache deutsch		Migrantenkinder		Alle Kinder	
	n <sup>a</sup>	Auffällig n (Anteil in %)	n <sup>a</sup>	Auffällig n (Anteil in %)	n <sup>a</sup>	Auffällig n (Anteil in %)
Spontansprache	170	9 (5,3)	73	6 (8,2)	253	16 (6,3)
Redefluss	169	5 (3,0)	73	5 (6,8)	246	10 (4,1)
Stimme	168	1 (0,6)	72	0 (0)	244	1 (0,4)
Sprachverständnis	170	4 (2,4)	73	14 (19,2)	253	19 (7,5)
Sprachproduktion	150	13 (8,7)	71	11 (15,5)	230	25 (10,9)
Wortschatz <sup>b</sup>	170	5 (2,9)	73	20 (27,4)	243	25 (10,3)
Grammatik <sup>b</sup>	169	5 (3,0)	73	24 (32,9)	242	29 (12,0)
Aussprache insgesamt	172	34 (19,8)	74	16 (21,6)	257	54 (21,0)
s-Laut	170	4 (2,4)	72	3 (4,2)	252	7 (2,8)
sch-Laut	169	23 (13,6)	72	8 (11,1)	251	32 (12,7)
Zischlaute insgesamt	169	25 (14,8)	71	8 (11,3)	250	34 (13,6)
r-Laut	171	8 (4,7)	73	5 (6,8)	255	14 (5,5)
g,k-Laute	169	7 (4,1)	72	1 (1,4)	251	9 (3,6)

<sup>a</sup> n kann wegen fehlender Daten oder unklaren Migrationshintergrunds variieren; <sup>b</sup> unterschiedliche Auffälligkeitskriterien für muttersprachlich deutsche und Kinder mit Migrationshintergrund.

**Tab. 3** Übereinstimmung zwischen sprachpädagogischer Auffälligkeit bzw. inklusiver Gesamtauffälligkeit im KiSS und Experten-Majoritätsurteil (freie Grundrate) über sprachpädagogische Auffälligkeit

Auffälligkeit im KiSS		Experten-Majoritätsurteil: sprachpädagogische Auffälligkeit		
		Auffällig n (Anteil in %)	Unauffällig n (Anteil in %)	Gesamt n (Anteil in %)
Sprachpädagogische Auffälligkeit	Auffällig	39 (16,0)	24 (9,9)	63 (25,9)
	Unauffällig	39 (16,0)	141 (58,0)	180 (74,1)
	Gesamt	78 (32,1) <sup>a</sup>	165 (67,9)	243 (100)
Inklusive Gesamtauffälligkeit	Auffällig	43 (17,9)	34 (14,2)	77 (32,1)
	Unauffällig	31 (12,9)	132 (55,0)	163 (67,9)
	Gesamt	74 (30,8)	166 (69,2)	240 (100)

<sup>a</sup> Diskrepanzen zwischen Zeilen- und Spaltensummen durch Rundungsfehler

klinischer Auffälligkeit zeigten zumeist, nämlich in 24 Fällen, auch eine sprachpädagogische Auffälligkeit, in den übrigen 11 Fällen nicht, z. B. bei Redefluss- und Stimmstörungen.

Aus Tab. 2 sind die Auffälligkeitsraten für die Untertests des KiSS bzw. der Hinweise auf bestimmte Sprech-, Sprach-, Stimm- und Redeflussstörungen zu entnehmen.

**Tab. 4** Übereinstimmung zwischen inklusiver klinischer Auffälligkeit bzw. inklusiver Gesamtauffälligkeit im KiSS und Experten-Majoritätsurteil (freie Grundrate) über inklusive klinische Auffälligkeit

Auffälligkeit im KiSS		Experten-Majoritätsurteil: inklusive klinische Auffälligkeit		
		Auffällig n (Anteil in %)	Unauffällig n (Anteil in %)	Gesamt n (Anteil in %)
Inklusive klinische Auffälligkeit	Auffällig	19 (7,9)	15 (6,2)	34 (14,2)
	Unauffällig	39 (16,3)	167 (69,6)	206 (85,8)
	Gesamt	58 (24,2)	182 (75,8)	240 (100)
Inklusive Gesamt- auffälligkeit	Auffällig	34 (14,2)	40 (16,7)	74 (30,9)
	Unauffällig	24 (10,0)	142 (59,2)	166 (69,2)
	Gesamt	58 (24,2)	182 (75,8)	240 (100)

**Tab. 5** Übereinstimmung zwischen exklusiver klinischer bzw. exklusiver Gesamtauffälligkeit im KiSS und Experten-Majoritätsurteil (freie Grundrate) über exklusive klinische Auffälligkeit

Auffälligkeit im KiSS		Experten-Majoritätsurteil: exklusive klinische Auffälligkeit		
		Auffällig n (Anteil in %)	Unauffällig n (Anteil in %)	Gesamt n (Anteil in %)
Exklusive klinische Auffälligkeit	Auffällig	18 (7,4)	9 (3,7)	27 (11,1)
	Unauffällig	39 (16,0)	177 (72,8)	216 (88,9)
	Normal	57 (23,5)	186 (76,5)	243 (100)
Exklusive Gesamt- auffälligkeit	Auffällig	52 (21,4)	16 (6,6)	68 (28,0)
	Unauffällig	44 (18,1)	131 (53,9)	175 (72,0)
	Gesamt	96 (39,5)	147 (60,5)	243 (100)

**Tab. 6** Übereinstimmung zwischen inklusiver Gesamtauffälligkeit im KiSS und im Experten-Majoritätsurteil (freie Grundrate)

Auffälligkeit im KiSS		Experten-Majoritätsurteil: inklusive Gesamtauffälligkeit		
		Auffällig n (Anteil in %)	Unauffällig n (Anteil in %)	Gesamt n (Anteil in %)
Inklusive Gesamt- auffälligkeit	Auffällig	49 (20,4)	25 (10,4)	74 (30,8)
	Unauffällig	43 (17,9)	123 (51,3)	166 (69,2)
	Gesamt	92 (38,3)	148 (61,7)	240 (100)

**Tab. 7** Übereinstimmung zwischen exklusiver Gesamtauffälligkeit im KiSS bzw. Auffälligkeit im HASE sowie SSV und dem Experten-Majoritätsurteil (freie Grundrate) über exklusive Gesamtauffälligkeit

Screening		Experten-Majoritätsurteil: exklusive Gesamtauffälligkeit		
		Auffällig n (Anteil in %)	Unauffällig n (Anteil in %)	Gesamt n (Anteil in %)
KiSS exklusive Gesamt- auffälligkeit	Auffällig	37 (15,2)	31 (12,8)	68 (28,0)
	Unauffällig	20 (8,2)	155 (63,8)	175 (72,0)
	Gesamt	57 (23,4)	186 (76,6)	243 (100)
HASE Auffälligkeit	Auffällig	47 (18,8)	24 (9,6)	71 (28,4)
	Unauffällig	49 (19,6)	130 (52,0)	179 (71,6)
	Gesamt	96 (38,4)	154 (61,6)	250 (100)
SSV Auffälligkeit	Auffällig	64 (25,9)	22 (8,9)	86 (34,8)
	Unauffällig	30 (12,1)	131 (53,0)	161 (65,2)
	Gesamt	94 (38,1)	153 (61,9)	247 (100)

## Zusammenhang zwischen KiSS-Befunden und Referenzverfahren

### Konfundierung der KiSS-Befunde mit dem Expertenurteil mit freier Grundrate

Für die verschiedenen Störungsbilder (sprachpädagogische Auffälligkeit, inklusive und exklusive klinische und Gesamtauffälligkeit) wurden die KiSS-Befunde mit dem jeweiligen Majoritätsurteil der Expertinnen konfundiert (kreuztabelliert; **Tab. 3, 4, 5, 6, 7**). Diese Vergleiche lagen nahe, da für die Einführung eines Sprachscreenings sowohl der Sprachförderbedarf (pädagogische Maßnahme) als auch der klinische Abklärungs- und Therapiebedarf (medizinische Maßnahme) von vorrangigem gesundheits- und bildungspolitischen Interesse sind. In die Darstellungen der klinischen Auffälligkeiten und Gesamtauffälligkeiten wurden einerseits Redefluss- und Stimmstörungen einbezogen (**Tab. 3, 4, 6**), da es sich dabei um therapiebedürftige Kommunikationsstörungen handelt. Andererseits (**Tab. 5, 7**) wurden sie in manchen Darstellungen weggelassen, da die exklusiven klinischen Auffälligkeiten im KiSS in guter Näherung die schriftspracherwerbsrelevanten SES wiedergeben und damit ein Vergleich mit anderen Screenings möglich ist.

Da zudem die Trennung in klinisch und sprachpädagogisch Auffällige methodisch schwierig erscheint und die Findung der Referenzkriterien das Ergebnis eines Optimierungsprozesses ist, gibt die Kreuztabellierung zwischen dem Expertenurteil zu sprachpädagogischen oder klinischen Auffälligkeiten und den Gesamtauffälligkeiten im KiSS die Validität des Screenings derzeit besser wieder als die Kreuztabellierung mit den sprachpädagogischen oder klinischen Auffälligkeiten im KiSS allein. Zumindest die medizinisch abklärungsbedürftigen Kinder sollten bei den Gesamtauffälligen weitestgehend eingeschlossen sein.

Die Screeninggütekriterien des KiSS gemessen am Experten-Majoritätsurteil mit freier Grundrate für die verschiedenen Konfundierungen finden sich in **Tab. 8**, ebenso die entsprechenden Werte für HASE und SSV. Die höchst-



**Tab. 8 Gütekriterien von KiSS, HASE und SSV, gemessen am Experten-Majoritätsurteil (freie Grundrate)**

Gütekriterium	Konfundierung: Auffälligkeit im KiSS, HASE und SSV vs. Auffälligkeit im Experten-Majoritätsurteil bei freier Grundrate									
	<i>Sprachpädagogisch auffällig: KiSS vs. Experten</i>	<i>Inklusiv gesamt-auffällig/KiSS vs. sprachpädagogisch auffällig/Experten</i>	<i>Inklusiv klinisch auffällig: KiSS vs. Experten</i>	<i>Inklusiv gesamt-auffällig/KiSS vs. klinisch auffällig/Experten</i>	<i>Exklusiv klinisch auffällig: KiSS vs. Experten</i>	<i>Exklusiv gesamt-auffällig/KiSS vs. exklusiv klinisch auffällig/Experten</i>	<i>Inklusiv gesamt-auffällig: KiSS vs. Experten</i>	<i>Exklusiv gesamt-auffällig: KiSS vs. Experten</i>	<i>Auffällig/HASE vs. gesamt-auffällig/Experten</i>	<i>Auffällig/SSV vs. gesamt-auffällig/Experten</i>
	<i>(95%-Konfidenzintervall)</i>									
Sensitivität (%)	50,0 (38,5–61,5)	55,8 (44,1–67,2)	32,8 (21,0–46,3)	58,6 (44,9–71,4)	31,6 (19,9–45,2)	64,9 (51,1–77,1)	53,3 (42,6–63,7)	54,2 (43,7–64,4)	49,0 (38,6–59,4)	68,1 (57,7–77,3)
Spezifität (%)	85,5 (79,1–90,5)	81,0 (74,1–86,7)	91,8 (86,8–95,3)	78,0 (71,3–83,8)	95,2 (91,0–97,8)	83,3 (77,2–88,4)	83,1 (76,1–88,8)	89,1 (82,9–93,7)	84,4 (77,7–89,8)	85,6 (79,0–90,8)
Rate falsch-positiv (%)	14,6 (9,6–20,9)	19,0 (13,3–25,9)	8,2 (4,7–13,2)	22,0 (16,2–28,7)	4,8 (2,2–9,0)	16,7 (11,6–22,8)	16,9 (11,2–23,9)	10,9 (6,4–17,1)	15,6 (10,3–22,3)	14,4 (9,2–21,0)
Rate falsch-negativ (%)	50,0 (38,5–61,5)	44,2 (32,8–55,9)	67,2 (53,7–79,0)	41,4 (28,6–55,1)	68,4 (54,8–80,1)	35,1 (22,9–48,9)	46,7 (36,3–57,4)	45,8 (35,6–56,3)	51,0 (40,6–61,4)	31,9 (22,7–42,3)
Positiver Vorhersagewert (%)	61,9 (48,8–73,9)	58,1 (46,1–69,5)	55,9 (37,9–72,8)	46,0 (34,3–57,9)	66,7 (46,0–83,5)	54,4 (41,9–66,6)	66,2 (54,3–76,8)	76,5 (64,6–85,9)	66,2 (54,0–77,0)	74,4 (63,9–83,2)
Negativer Vorhersagewert (%)	78,3 (71,6–84,1)	79,5 (72,6–85,4)	81,1 (75,0–86,2)	85,5 (79,3–90,5)	81,9 (76,2–86,8)	88,6 (82,9–92,9)	74,1 (66,7–80,6)	74,9 (67,8–81,1)	72,6 (65,5–79,0)	81,4 (74,5–87,1)
Effizienz	74,1 (68,1–79,5)	72,9 (66,8–78,4)	77,5 (71,7–82,6)	73,3 (67,3–78,8)	80,3 (74,7–85,1)	79,0 (73,4–84,0)	71,7 (65,5–77,3)	75,3 (69,4–80,6)	70,8 (64,7–76,4)	79,0 (73,3–83,9)
Youden-Index	35,5 (23,1–47,8)	36,8 (24,2–49,5)	24,5 (11,8–37,2)	36,6 (22,6–50,7)	26,7 (14,3–39,2)	48,3 (34,8–61,7)	36,4 (24,5–48,2)	43,3 (32,1–54,5)	33,4 (21,9–44,9)	53,7 (42,8–64,7)
Positive Likelihood Ratio	3,44	2,94	3,98	2,67	6,53	3,90	3,15	4,98	3,14	4,74
Negative Likelihood Ratio	0,59	0,55	0,73	0,53	0,72	0,42	0,56	0,51	0,61	0,37
Kontingenzkoeffizient (phi, $\phi$ )	0,38	0,37	0,30	0,34	0,36	0,46	0,38	0,47	0,36	0,55
RATZ-Index	0,43	0,38	0,42	0,40	0,56	0,51	0,45	0,61	0,45	0,59

Positive Likelihood Ratio: >10 sehr gut, >3 gut; negative Likelihood Ratio: <0,1 sehr gut, <0,3 gut.

te Spezifität (95%) wurde in der Konfundierung der exklusiv klinisch im KiSS und im Experten-Majoritätsurteil Auffälligen erzielt, die höchste Sensitivität (65%) in der Konfundierung der im KiSS exklusiv Gesamtauffälligen mit den exklusiv klinisch Auffälligen im Expertenurteil. Wurde sprachpädagogische oder klinische Auffälligkeit (exklusiv oder inklusiv) im KiSS mit dem freien Expertenurteil der entsprechenden Kategorie konfundiert, zeigten sich gute Spezifitäten von 86–95% (■ **Tab. 8**), aber wenig befriedigende Sensitivitäten (32–50%). Wurde hingegen überprüft, ob sich die von den Experten als sprachpädagogisch oder klinisch auffällig klassifizierten Kinder zumindest unter den im KiSS Gesamtauffälligen wiederfanden, stieg die Sensitivität auf akzeptablere, aber gleichwohl noch nicht voll befriedigende Werte (56–65%), bei erwartungsgemäßer Minderung der Spezifität (78–83%).

**Tab. 9 Übereinstimmung zwischen inklusiver klinischer bzw. Gesamtauffälligkeit im KiSS und Experten-Majoritätsurteil bzgl. der 34 inklusiv klinisch auffälligsten Kinder**

Auffälligkeit im KiSS	Experten-Majoritätsurteil: 34 inklusiv klinisch Auffälligste			
	<i>Auffällig n (Anteil in %)</i>	<i>Unauffällig n (Anteil in %)</i>	<i>Gesamt n (Anteil in %)</i>	
Inklusive klinische Auffälligkeit	Auffällig	20 (8,3)	14 (5,8)	34 (14,2)
	Unauffällig	13 (5,4)	193 (80,4)	206 (85,8)
	Gesamt	33 (13,8)	207 (86,3)	240 (100)
Inklusive Gesamtauffälligkeit	Auffällig	29 (12,1)	45 (18,8)	74 (30,8)
	Unauffällig	4 (1,7)	162 (67,5)	166 (69,2)
	Gesamt	33 (13,8)	207 (86,3)	240 (100)

### Konfundierung der KiSS-Befunde mit dem Majoritäts- bzw. Konsens-Expertenurteil bei festgelegter Grundrate

Um die freien Expertinnenurteile mit ihren zu hohen und zu stark variierenden Grundraten zu optimieren, wurden prävalenzbasierte Grundraten vorgegeben, in der Hoffnung, hierdurch eine bessere Übereinstimmung zwischen KiSS-Befunden und Expertenurteilen zu erzielen,

speziell für die unbedingt zu identifizierenden klinisch auffälligen Kinder (SES, Redefluss- und Stimmstörungen). In ■ **Tab. 9 und 10** sind die im KiSS inklusiv oder exklusiv klinisch oder Gesamtauffälligen dargestellt. Diese sind hier zum einen mit dem Konsens-Expertenurteil bezüglich der 34 inklusiv klinisch auffälligsten Kinder konfundiert und zum anderen mit dem Experten-Majoritätsurteil zu den 27 exklusiv klinisch auffälligsten

**Tab. 10** Übereinstimmung zwischen exklusiver klinischer bzw. inklusiver sowie inklusiver Gesamtauffälligkeit im KiSS und Experten-Majoritätsurteil bzgl. der 27 exklusiv klinisch auffälligsten Kinder

Auffälligkeit im KiSS		Experten-Majoritätsurteil: 27 exklusiv klinisch Auffälligste		
		Auffällig n (Anteil in %)	Unauffällig n (Anteil in %)	Gesamt n (Anteil in %)
Exklusive klinische Auffälligkeit	Auffällig	12 (4,9)	15 (6,2)	27 (11,1)
	Unauffällig	14 (5,8)	202 (83,1)	216 (88,9)
	Gesamt	26 (10,7)	217 (89,3)	243 (100)
Exklusive Gesamt- auffälligkeit	Auffällig	21 (8,7)	47 (19,3)	68 (28,0)
	Unauffällig	5 (2,1)	170 (70,0)	175 (72,0)
	Gesamt	26 (10,7)	217 (89,3)	243 (100)
Inklusive Gesamt- auffälligkeit	Auffällig	21 (8,7)	53 (22,1)	74 (30,8)
	Unauffällig	4 (1,7)	162 (67,5)	166 (69,2)
	Gesamt	25 (10,4)	215 (89,6)	240 (100)

Kindern (nicht für sprachpädagogische auffällige Kinder gültig, da es für Sprachförderbedarf keine verlässlichen Prävalenzangaben gibt). Die entsprechenden Gütekriterien sind in **Tab. 11** gezeigt.

Bei diesen Referenzkriterien lieferte die Konfundierung der 34 im KiSS inklusiv klinisch Auffälligen mit dem entsprechenden Konsens-Expertenurteil nicht nur die höchste Spezifität (93%) und eine akzeptable Sensitivität (61%), sondern auch die höchste Effizienz (Anteil richtiger an allen Entscheidungen, 89%). Die höchste Sensitivität von 88% ergab die Konfundierung der im KiSS identifizierten Fälle von inklusiver Gesamtauffälligkeit mit dem Konsens-Expertenurteil über die 34 inklusiv klinisch auffälligsten Kinder, was eine Rate von 12% im KiSS falsch-negativen Bewertungen bedeutet. Diese Ergebnisse belegen, dass eine angemessene Beschränkung der professionellen Sprachdiagnostikerinnen auf den Bereich der zu erwartenden Prävalenzraten zu einer deutlichen Verbesserung der Gütekriterien des KiSS führt, und zwar allein aufgrund der Schärfung des Referenzkriteriums.

Von den 34 im KiSS inklusiv klinisch auffälligen Kindern wurden durch pädiatrische und weiterführende Untersuchungen und die durch die Fragebögen erhobenen anamnestischen Befunde Komorbiditäten in 5 Fällen festgestellt (2 Intelligenzminderungen, 3 tiefgreifende Entwicklungsstörungen). Unter den im Experten-Majoritätsurteil als SES-verdächtig klassifizierten Fällen fanden sich 2 Kinder mit Komorbiditäten. Mit diesem Referenzkriterium hätten etwa 93% der SES-Kin-

der eine UESS und etwa 7% eine SES assoziiert mit einer Komorbidität. Das ergibt Prävalenzraten von 10,5% für UESS und 0,8% für SES assoziiert mit einer Komorbidität in der Gesamtstichprobe.

### Vergleich der Screenings HASE, SSV und KiSS

Das HASE dient der Risikoabklärung für bestimmte sprachliche und schriftsprachliche Bereiche, d. h. der Früherkennung von Lese- und Rechtsschreibstörungen und der Auffindung von Sprach- und Schriftspracherwerbsstörungen [12]. Das SSV zielt auf die Identifikation von Kindern mit dem Risiko einer Sprachentwicklungsverzögerung ab [25]. Gemeint ist allerdings ein Sprachdefizit, das entweder einer Sprachförderung bedarf oder einer Therapie i. S. einer SES. Damit entsprechen die Auffälligen in SSV und HASE am ehesten dem exklusiven Expertenurteil zur Gesamtauffälligkeit im KiSS bei freier Grundrate. KiSS und SSV erreichen gleichwertige Gütekriterien, einschließlich RATZ-Index, HASE etwas darunter liegende, wobei dieses Screening jedoch für ältere Kinder (5- bis 6-Jährige) konstruiert wurde.

### Diskussion

Obligatorische Kriterien für Studien zur Bewertung der Testgüte eines universellen Screenings sind ein unselektiertes Screeningkollektiv, eine geeignete Referenzmethode, das Screening des gesamten Kollektivs, die Untersuchung aller Kinder mit dem Referenztest, lückenlose Stich-

probendokumentation und Ergebnisdarstellung in 4-Felder-Tafeln [28]. Bis auf die eingeschränkte, aber akzeptable Eignung der (einzig verfügbaren) Referenzmethode wurden alle diese Kriterien in der vorliegenden Studie erfüllt.

Die Studie sollte die konkurrente klassifikatorische Validität des KiSS mittels geeigneter Referenzkriterien belegen. Dabei war zunächst aus den verfügbaren deutschen Sprachtests ein bestmögliches Referenzkriterium zu erstellen. Von mehreren Vorgehensweisen hat sich das gemeinsame Expertenurteil mit vorgegebener Grundrate am besten bewährt. Damit wurden die in amerikanischen Studien ermittelten Richtwerte erreicht, d. h. gute Sensitivitäten zwischen 80 und 88% und sehr gute RATZ-Indizes zwischen 73 und 82%, allerdings nur dann, wenn die exklusive oder inklusive Gesamtauffälligkeit im KiSS mit dem Expertenurteil zu exklusiver (SES) oder inklusiver klinischer Auffälligkeit (SES, Redefluss- und Stimmstörungen) kreuztabelliert wurde. Die Sensitivitäten waren weniger hoch, wenn exklusive bzw. inklusive klinische oder sprachpädagogische Auffälligkeit im KiSS mit dem direkt vergleichbaren Expertenurteil kreuztabelliert wurde. Hohe Sensitivitäten wie die genannten gehen allerdings zu Lasten der Spezifitäten, die bei 75% bzw. 78% lagen.

Zusammenfassend lässt sich festhalten, dass diese Werte für ein Sprachscreening durchaus akzeptabel sind und das KiSS als wertvolles Werkzeug ausweisen. In den anderen Kreuztabellierungen (klinische oder sprachpädagogische Auffälligkeit im KiSS mit direkt vergleichbarem Expertenurteil, **Tab. 8 und 11**) lagen die Spezifitäten zwischen 85 und 95% und waren damit sehr gut bis gut. Für sprachpädagogische Auffälligkeit konnten wegen fehlender Referenzkriterien den Expertinnen keine Grundraten vorgegeben werden, was sich auf die diesbezüglichen Gütekriterien des KiSS ungünstig auswirken musste, aber nicht gegen dessen Qualität spricht. Somit werden durch das KiSS mit hoher Wahrscheinlichkeit die klinisch abzuklärenden Kinder identifiziert (Rate Falsch-Negativer 12%). Für die Identifikation sprachpädagogisch Auffälliger fehlt ein zureichend verlässliches Referenzkriterium, solange die prädiktive Validität des

**Tab. 11** Gütekriterien des KiSS, gemessen am Experten-Majoritätsurteil mit festgelegten Grundraten

Gütekriterium	Konfundierung: Auffälligkeit im KiSS vs. Auffälligkeit im Experten-Majoritätsurteil bei fester Grundrate				
	Inklusive klinische Auffälligkeit vs. 34 inklusiv klinisch Auffälligste (95%-Konfidenzintervall)	Inklusive Gesamtauf-fälligkeit vs. 34 inklusiv klinisch Auffälligste (95%-Konfidenzintervall)	Exklusive klinische Auffälligkeit vs. 27 exklusiv klinisch Auffälligste	Exklusive Gesamtauf-fälligkeit vs. 27 exklusiv klinisch Auffälligste	Inklusive Gesamtauf-fälligkeit vs. 27 exklusiv klinisch Auffälligste
Sensitivität (%)	60,6 (42,1–77,1)	87,9 (71,8–96,6)	46,2 (26,6–66,6)	80,8 (60,7–93,5)	84,0 (63,9–95,5)
Spezifität (%)	93,2 (88,9–96,3)	78,3 (72,0–83,7)	93,1 (88,9–96,1)	78,3 (72,3–83,6)	75,4 (69,0–81,0)
Rate falsch-positiv (%)	6,8 (3,8–11,1)	21,7 (16,3–28,0)	6,9 (3,9–11,1)	21,7 (16,4–27,7)	24,7 (19,0–31,0)
Rate falsch-negativ (%)	39,4 (22,9–57,9)	12,1 (3,4–28,2)	53,9 (33,4–73,4)	19,2 (6,6–39,4)	16,0 (4,5–36,1)
Positiver Vorhersagewert (%)	58,8 (40,7–75,4)	39,2 (28,0–51,2)	44,4 (25,5–64,7)	30,9 (20,2–43,3)	28,4 (18,5–40,1)
Negativer Vorhersagewert (%)	93,7 (89,5–96,6)	97,6 (94,0–99,3)	93,5 (89,4–96,4)	97,1 (93,5–99,1)	97,6 (94,0–99,3)
Effizienz	88,8 (84,1–92,5)	79,6 (73,9–84,5)	88,1 (83,3–91,9)	77,6 (72,9–83,6)	76,3 (70,4–81,5)
Youden-Index	53,8 (36,8–70,9)	66,1 (53,7–78,6)	39,2 (19,8–58,7)	56,1 (43,0–75,2)	59,4 (43,9–74,8)
Positive Likelihood Ratio	8,96	4,04	6,68	3,73	3,41
Negative Likelihood Ratio	0,42	0,16	0,58	0,25	0,21
Kontingenzkoeffizient ( $\phi$ )	0,53	0,49	0,39	0,41	0,39
RATZ-Index	0,48	0,82	0,39	0,73	0,81

KiSS für den Schriftspracherwerb nicht bekannt ist. Derzeit ist die Unterscheidung zwischen klinischer und sprachpädagogischer Auffälligkeit im Screening noch nicht befriedigend gelöst und bedarf einer Optimierung. Die entsprechenden RATZ-Indizes erreichten übrigens in allen Konfundierungen (■ **Tab. 8 und 11**) Werte, die für Sprachstandsscreenings als mindestens gut angesehen werden.

Die Qualität des KiSS wie auch anderer Sprachscreenings ohne belegte prädiktive Validität kann nur eingeschränkt beurteilt werden, solange die Gütekriterien des konkurrierenden externen Referenzkriteriums selbst, also das Profil aus den einzelnen Referenztests, für eine Auffälligkeitsdiagnose nicht nachgewiesen sind und zu Zweifel Anlass geben. Das trifft insbesondere auf die subjektive klinische Beurteilung von Sprachtests durch Experten zu. Da eine solche Bewertung täglich in der praktischen Arbeit von Sprachtherapeuten vorgenommen wird, sind mangelnde Untersucherübereinstimmung und Überdiagnose möglicherweise eher Regel als Ausnahme. Das Problem ist jedoch international bekannt und liegt (auch) in der Natur der Sprache mit ihren verschiedenen linguistischen Ebenen und deren Entwicklungsvarianzen.

Befriedigende Referenzkriterien stellen auch anderorts eine der Hauptschwierigkeiten bei der Bewertung eines Sprachscreenings dar. Gescheitert sind Versuche, aus den kontinuierlichen Verteilun-

gen der Sprachtestdaten mit taxometrischen Analysen eine kategoriale Sprachentwicklungsstörung abzuleiten, also Kinder zu identifizieren, die sich nicht nur im Grad der Sprachkompetenz, sondern in der Art von anderen Kindern unterscheiden [17]. Deswegen gilt häufig als Auffälligkeitskriterium das Unterschreiten der 1.–2. Standardabweichung unterhalb des Mittelwerts von Sprachtests [43]. Werden mehrere Sprachtests durchgeführt, erfolgt die Profildbewertung nach subjektiven Kriterien, mit erheblichen Beeinträchtigungen von Auswerterobjektivität und Reliabilität. In einer US-amerikanischen Studie der National Institutes of Health ließen Records u. Tomblin [45] 27 Sprachexperten die Sprachprofile von 92 Kindern anhand von 4 Sprachtests und dem Wechsler-Intelligenztest auf das Vorliegen von Sprachstörungen beurteilen. Die Beurteilerübereinstimmung war völlig unbefriedigend (generalisiertes  $\kappa=0,14$ ), und selbst die Intraraterreliabilität für wiederholte Beurteilungen unerkannt untergeschobener Befunddubletten war mangelhaft ( $\phi=0,68$ ). Die Datenanalyse zeigte, dass kein Sprachexperte alle 5 Tests für das Auffälligkeitsurteil verwendete, die meisten von ihnen benutzten nur 2 oder 3. Die Diagnostiker gingen also nach subjektiven, impliziten Algorithmen vor.

Diagnostische Urteile mit solch unbefriedigenden Objektivitäten und Retestreliabilitäten können nicht zureichend valide sein und verdienen nicht das Referenz-

prädikat „Goldstandard“, sondern höchstens das Prädikat „Katzengoldstandard“. (Volkstümlich wird der golden glänzende Pyrit auch Katzensgold genannt.) Der Testgüte eines Sprachscreenings sind folglich deutlich reduzierte Obergrenzen gesetzt. Dennoch werden weltweit Sprachscreenings und sprachdiagnostische Tests notwendigerweise eingesetzt, unter Inkaufnahme ihrer Beschränkungen. Bei den punkt-biserialen Korrelationen zwischen den verschiedenen Auffälligkeitsurteilen der einzelnen Expertinnen und den Kennwerten der Referenztests zeigten sich auch in unserer Studie die Auswirkungen subjektiver Auffälligkeitsalgorithmen. Die Sprachheilpädagogin und die Linguistin erreichten durchweg höhere Korrelationen als die Logopädin mit Testleistungen in Sprachverständnis, Wortschatz und Grammatik, die Logopädin hingegen zeigte höhere Korrelationen mit den Testwerten in der Aussprache.

Eine weitere Schwierigkeit bestand in der Anforderung an das hier beschriebene Screening, nicht nur die übliche dichotome Entscheidung zu treffen, sondern eine trichotome (klinisch auffällig/sprachpädagogisch auffällig/unauffällig). Eine derartige Dreiteilung wird auch von keinem anderen uns bekannten Screeningverfahren vorgenommen. Sie erscheint aber aus gesellschafts- und gesundheitspolitischer Perspektive geboten, da das Bild der Spracherwerbsdefizite komplex ist und sehr verschieden gelagerte, durch unter-

schiedliche Berufsgruppen zu vertretende Konsequenzen nach sich zieht. SES, Redefluss- und Stimmstörungen bedürfen einer Therapie, sprachliche Defizite anderer Ursache einer Sprachförderung. Da selbst für erfahrene Sprachdiagnostiker eine solche Entscheidung häufig schwierig und nur aus dem Verlauf des Störungsbilds heraus zu treffen ist, war anzunehmen, dass eine hinreichend valide Dreiteilung nur Ergebnis eines diffizilen Optimierungsprozesses sein könnte.

Sollen allerdings die unbedingt zu identifizierenden Kinder mit SES, Redefluss- und/oder Stimmstörungen detektiert werden, dann geschieht das durch das KiSS entweder sofort (Kind klinisch auffällig), oder das Kind gelangt mit hoher Restwahrscheinlichkeit in die Gruppe der sprachpädagogisch zu fördernden Kinder, wo eine erhöhte Sensibilisierung und Kenntnis der fördernden Erzieherinnen für Sprachstörungen angenommen werden darf. Ein solches Kind würde bei ausbleibendem Förderfortschritt regulär nach einigen Wochen eine Nachtestung durch eine Sprachexpertin erhalten und dann doch in die medizinische Abklärung gelangen. Entsprechend würde die Mehrzahl der nach dem Expertenurteil förderbedürftigen Kinder im KiSS als sprachpädagogisch auffällig oder als rein klinisch auffällig klassifiziert werden. Im selteneren letzteren Fall sollte die Diagnostik dann den Sprachförderbedarf aufdecken. Für die übrigen, nur von den Expertinnen als sprachpädagogisch auffällig klassifizierten Kinder sei auf die aufgeführte Problematik des fehlenden objektiveren Referenzkriteriums für Sprachförderbedarf hingewiesen.

Das KiSS wird weiter optimiert. Derzeit werden elektrophysiologische Untersuchungen durchgeführt, um ggf. Trennungsalgorithmen zwischen klinischen (sprachsystemrelevanten) und leichteren Sprachauffälligkeiten auf besser objektivierbare Parameter zu gründen. Verglichen mit etablierten Sprachscreenings, die auf Nachsprechleistungen von grammatischen und phonologischen Strukturen basieren (HASE, SSV), schnitt das KiSS bereits in der konkurrenten Kriteriumsvalidität numerisch etwas besser ab [19] und in der hier beschriebenen kategorialen Bewertung sprachlicher Auffäl-

igkeiten mindestens gleichwertig. Allerdings liegen für das KiSS noch keine prädiktiven Validitätsmaße für den Schriftspracherwerb vor, die für das HASE und weitere vergleichbare Nachsprechtests belegt sind [12]. Um eine angemessene prädiktive Validität von vornherein zu sichern, wird der Test derzeit um 2 Untertests (Kunstwörter und Sätze nachsprechen) erweitert.

Ein Screening nur auf UESS erscheint kaum möglich, da Komorbiditäten auch verborgen sein können. Vielmehr ist ein Screening auf SES, in denen UESS enthalten sind, sinnvoll und möglich, so auch mit dem KiSS. Einige Komorbiditäten sind offensichtlich, andere hingegen bleiben verborgen, z. B. geringgradige Hörstörungen. Ein Sprachscreening könnte zwar die SES aufdecken, nicht aber die Komorbidität. Hier gäbe erst eine weiterführende Diagnostik Aufschluss, ob eine UESS oder eine SES mit Komorbidität vorliegt. Die von uns ermittelte Komorbiditätsrate von 7% unter allen SES-Fällen weist auf den relativ hohen Anteil dieser Störungskombinationen in Regelkindergärten hin. Auch ein Screening auf Sprachauffälligkeiten insgesamt erscheint sinnvoll, solange das weitere Prozedere mit sequenzieller oder simultaner Zuordnung zu Sprachförder- oder medizinischem Diagnostikbedarf algorithmisch definiert ist.

### Fazit für die Praxis

**Zusammenfassend liegt mit dem KiSS ein Sprachscreening für vier- bis viereinhalbjährige Kinder vor, das zuverlässig und mit ausreichender konkurrenter und Konstruktvalidität Kinder mit Sprachentwicklungsstörungen identifiziert, darunter auch UESS, Redefluss- und Stimmstörungen. Das Verfahren kann möglicherweise auch Kinder mit einem reinen Sprachförderbedarf mit akzeptabler Validität detektieren. Hierfür liegt aber kein ausreichend verlässliches Referenzkriterium vor. Auch aus diesem Grunde bleibt die Machbarkeit einer Trennung zwischen Sprachförderbedarf und medizinischem Abklärungsbedarf ungeklärt und u. a. deswegen das KiSS optimierungsbedürftig. Mit erzwungener Schärfung der Referenzkriterien durch die Vorgabe prävalenzbasierter Grundraten ver-**

**bessern sich die Validitätswerte des KiSS. Daher ist die Entwicklung eines deutschsprachigen Referenztests mit algorithmischen Entscheidungskriterien für die zeiteffiziente Überprüfung aller linguistischen Ebenen wünschenswert. Damit kann für die Bewertung von Sprachscreenings vielleicht kein Goldstandard geschaffen werden, der diesen Namen verdient. Die derzeit verwendete professionelle Kindersprachdiagnostik stellt sich im Licht strenger testmethodischer Gütekriterien eher als Katzensgoldstandard heraus, sodass ein solider „Messingstandard“ das Nahziel sein könnte.**

### Korrespondenzadresse

Prof. Dr. K. Neumann



Schwerpunkt für Phoniatrie und Pädaudiologie  
Klinik für Hals-Nasen-Ohren-Heilkunde, Goethe-Universität  
Theodor-Stern-Kai 7, Haus 7A,  
60590 Frankfurt am Main  
Katrin.Neumann@em.uni-frankfurt

**Danksagung.** Die Erstellung des Verfahrens wurde vom Hessischen Sozialministerium in Auftrag gegeben. Wir danken Andreas Bodensohn, Barbara Jung, Amelie Löffler, Johannes Mertens, Arkadius Neumann, Carla Sack und den Projektverantwortlichen beim Hessischen Sozialministerium, Dr. Angela Wirtz und Dr. Klaus Jahn, für deren Unterstützung und Rat.

**Interessenkonflikt.** Der korrespondierende Autor weist auf folgende Beziehung(en) hin: Gefördert durch das Hessische Sozialministerium V4.1-18h26.47.01.

### Literatur

1. Ackermann H (2008) BiAS für Windows, Version 8.6: Biometrische Analyse von Stichproben. Epsilon, Hochheim
2. American Psychiatric Association (2000) Diagnostic and statistical manual of mental disorders, (4. Aufl., text rev.) Author, Washington/DC
3. Aram DM, Ekelman BL, Nation JE (1984) Preschoolers with language disorders: 10 years later. J Speech Lang Hear Res 27:232–244
4. Aram DM, Nation JE (1980) Preschool language disorders and subsequent language and academic difficulties. J Commun Disord 13:159–170
5. Baddeley A (1992) Working memory. Science 255:556–559
6. Bashir AS, Scavuzzo A (1992) Children with language disorders: Natural history and academic success. J Learn Disabil 25:53–65
7. Beitchman JH, Wilson B, Brownlie EB et al (1996a) Longterm consistency in speech/language profiles: I. Developmental and academic outcomes. J Am Acad Child Adolesc Psychiatry 35:804–814

8. Beitchman JH, Wilson B, Brownlie EB et al (1996b) Longterm consistency in speech/language profiles: II. Behavioral, emotional and social outcomes. *J Am Acad Child Adolesc Psychiatry* 35:814–825
9. Bishop DVM, Laws G, Adams C, Norbury CF (2006) High heritability of speech and language impairments in 6-year-old twins demonstrated using parent and teacher report. *Behav Genet* 36:173–184
10. Bortz J, Lienert GA, Boehnke K (1990) Verteilungsfreie Methoden in der Biostatistik. Springer, Berlin
11. Botting N, Conti-Ramsden G (2000) Social and behavioural difficulties in children with specific language impairments. *Child Lang Teach Ther* 16:105–120
12. Brunner M, Schöler H (2002) HASE – Heidelberger Auditives Screening in der Einschulungsuntersuchung. Westra, Wertingen
13. Canning PM, Lyon ME (1989) Young Children with Special Needs: Prevalence and Implications in Nova Scotia. *Can J Educ* 14:368–380
14. Conti-Ramsden G, Durkin K, Simkin Z, Knox E (2008a) Specific language impairment and school outcomes. I: Identifying and explaining variability at the end of compulsory education. *Int J Lang Commun Disord* 7:1–21
15. Conti-Ramsden G, Durkin K, Simkin Z, Knox E (2008b) Specific language impairment and school outcomes. II: Educational context, student satisfaction, and post-compulsory progress. *Int J Lang Commun Disord* 7:1–20
16. Deutsches Institut für Medizinische Dokumentation (2005) Internationale Statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme. <http://www.dimdi.de/static/de/klasi/diagnosen/icd10/htmlgm2010/block-f80-f89.htm#F80>, Stand 10.04.10
17. Dollaghan CA (2004) Taxometric analyses of specific language impairment in 3- and 4-year-old children. *J Speech Lang Hear Res* 47:464–475
18. Durkin K, Wadman R, Conti-Ramsden G (2008) Self-esteem, shyness, and sociability in adolescents with specific language impairment (SLI). *J Speech Lang Hear Res* 51:938–952
19. Euler HA, Holler-Zittlau I, Minnen S von et al (2010) Kinder-Sprach-Screening (KiSS). Psychometrische Gütekriterien eines Kurztests zur Erfassung des Sprachstands 4-jähriger Kinder. *HNO* 58:116–123
20. Farrington DP, Loeber R (1989) Relative improvement over chance (RIOC) and phi as measures of predictive efficiency and strength of association in 2x2 tables. *J Quant Criminol* 5:201–213
21. Felsenfeld S, Broen PA, McGue M (1992) A 28-year-follow-up of adults with a history of moderate phonological disorder: linguistic and personality results. *J Speech Lang Hear Res* 35:1114–1125
22. Fleiss JL (1971) Measuring nominal scale agreement among many raters. *Psychol Bull* 76:378–382
23. Fox A (2005) PLAKSS – Psycholinguistische Analyse kindlicher Sprechstörungen, 2. Aufl. Harcourt Test Services, Frankfurt/Main
24. Green DM, Swets JM (1966) Signal detection theory and psychophysics. Wiley, New York/NY
25. Grimm H (2003) SSV – Sprachscreening für das Vorschulalter. Hogrefe, Göttingen
26. Hasselhorn M, Grube D (2003) Das Arbeitsgedächtnis. Funktionsweise, Entwicklung und Bedeutung für kognitive Leistungsstörungen. *Sprache Stimme Gehör* 27:31–37
27. Holler-Zittlau I, Dux W, Berger R (2003) Marburger Sprach-Screening für 4- bis 6-jährige Kinder (MSS). Persen, Nornburg/Niederelbe
28. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Hrsg) (2009) Abschlussbericht S06-01: Früherkennungsuntersuchung auf umschriebene Entwicklungsstörungen des Sprechens und der Sprache. Version 1.0; 17.06.2009. [http://www.iqwig.de/download/S06-01\\_Abschlussbericht\\_Fruherkennung\\_umschriebener\\_Stoerungen\\_des\\_Sprechens\\_und\\_der\\_Sprache.pdf](http://www.iqwig.de/download/S06-01_Abschlussbericht_Fruherkennung_umschriebener_Stoerungen_des_Sprechens_und_der_Sprache.pdf)
29. Kauschke C, Siegmüller J (2002) Pathologische Diagnostik bei Sprachentwicklungsstörungen. Urban und Fischer, München
30. Kiese-Himmel C (1999) Ein Jahrhundert Forschung zur gestörten Sprachentwicklung. *Sprache Stimme Gehör* 23:128–137
31. Kiese-Himmel C (2005) AWST-R. Aktiver Wortschatztest für 3- bis 5-jährige Kinder – Revision. Hogrefe, Göttingen
32. Kiese-Himmel C (2008) Entwicklung sprach- und kommunikationsgestörter Kinder, am Beispiel von „Late Talkers“ sowie Kindern mit spezifischen Sprachentwicklungsstörungen. In: Hasselhorn M, Silbereisen RK (Hrsg) Entwicklungspsychologie des Säuglings- und Kindesalters (Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung. Serie V: Entwicklungspsychologie, Bd 4). Hogrefe, Göttingen, S 693–730
33. Lattermann C, Euler HA, Neumann K (2008) A randomized control trial to investigate the impact of the Lidcombe 1 Program on early stuttering in German-speaking preschoolers. *J Fluency Disord* 33:52–65
34. Law J, Boyle J, Harris F et al (2000) The feasibility of universal screening for primary speech and language delay: findings from a systematic review of the literature. *Dev Med Child Neurol* 42:190–200
35. Law J, Garrett Z, Nye C (2004) The efficacy of treatment for children with developmental speech and language delay/disorder: a meta-analysis. *J Speech Lang Hear Res* 47:924–943
36. Law J, Harris F (2000) The parameters of early language screening. *Child Care Health Dev* 26:83–89
37. Leonard LB (2000) Specific language impairments across languages. In: Bishop DVM, Leonard LB (Hrsg) Speech and language impairments in children: causes, characteristics, intervention, and outcome. Taylor & Francis, Philadelphia, S 115–129
38. Marx H (1992) Methodische und inhaltliche Argumente für und wider eine frühe Identifikation und Prädiktion von Lese-Rechtschreibschwierigkeiten. *Diagnostica* 38:249–268
39. Marx H, Jansen H, Skowronek H (2000) Prognostische, differentielle und konkurrente Validität des Bielefelder Screenings zur Früherkennung von Lese-Rechtschreibschwierigkeiten. In: Hasselhorn M, Schneider W, Marx H (Hrsg) Diagnostik von Lese-Rechtschreibschwierigkeiten. Hogrefe, Göttingen, S 9–34
40. Nelson HD, Nygren P, Walker M, Panoscha R (2006) Screening for speech and language delay in preschool children: Systematic evidence review for the US Preventive Services Task Force. *Pediatrics* 117:e298–e319
41. Neumann K, Keilmann A, Kiese-Himmel C et al (2008) Leitlinien der Deutschen Gesellschaft für Phoniatrie und Pädaudiologie zu Sprachentwicklungsstörungen bei Kindern. <http://leitlinien.net/049-006.htm>, Stand 13.04.10
42. Neumann K, Keilmann A, Rosenfeld J et al (2009) Leitlinien der Deutschen Gesellschaft für Phoniatrie und Pädaudiologie (gekürzte Fassung). *Kindh Entwickl* 18:222–231
43. Paul R (2007) Language disorders from infancy through adolescence, 3. Aufl. Mosby Elsevier, St. Louis/MO
44. Plante E, Vance R (1994) Selection of preschool language tests: A data-based approach. *Lang Speech Hear Serv Sch* 25:15–24
45. Records NL, Tomblin JB (1994) Clinical decision making: Describing the decision rules of practicing speech-language pathologists. *J Speech Hear Res* 37:144–156
46. Reynell J (1997) The Reynell Developmental Language Scales III. NFER-Nelson, Windsor
47. Ribeiro A (2006) Funktionelle Stimmstörungen im Kindesalter. Schulz-Kirchner, Idstein
48. Schnieders G, Komor A (2007) Eine Synopse aktueller Verfahren der Sprachstandsfeststellung. In: Ehlich K (Hrsg) Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund. Bundesministerium für Bildung und Forschung (BMBF), Referat Öffentlichkeitsarbeit, Berlin, S 261–342
49. Snowling MJ, Bishop DVM, Stothard SE et al (2006) Psychosocial outcomes at 15 years of children with a preschool history of speech-language impairment. *J Child Psychol Psychiatry* 48:759–765
50. Thomson C, Polnay L (Hrsg) (2002). Community paediatrics, 3. Aufl. Elsevier, Edinburgh
51. Tomblin JB, Bruckwalter PR (1994) Studies of genetics of specific language impairment. In: Watkins RV, Rice ML (Hrsg) Specific language impairments in children. Brookes, Baltimore, S 17–34
52. Tomblin JB, Records NL, Buckwalter P et al (1997) Prevalence of specific language impairment in kindergarten children. *J Speech Lang Hear Res* 40:1245–1260
53. U.S. Preventive Services Task Force (2003) Guide to clinical preventive services, 3. Aufl. <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hstat3.section.34225>, Stand 15.11.2008
54. Suchodoletz W von (2003) Umschriebene Sprachentwicklungsstörungen. *Monatsschr Kinderheilkd* 151:31–37
55. Suchodoletz W von (2009) Wie wirksam ist Sprachtherapie? *Kindh Entwickl* 18:213–221
56. Yairi E, Ambrose NG (1999) Early childhood stuttering I: Persistence and recovery rates. *J Speech Lang Hear Res* 42:1097–1112